Genome-wide association study identifies TF as a significant modifier gene of iron metabolism in HFE hemochromatosis

Marie De Tayrac, Marie-Paule Roth, Anne-Marie Jouanolle, Hélène Coppin,
Gérald Le Gac, Alberto Piperno, Claude Férec, Sara Pelucchi, Virginie Scotet,
Edouard Bardou-Jacquet, et al.

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Genome-wide association study identifies *TF* as a significant modifier gene of iron metabolism in *HFE* hemochromatosis

Short Title – GWAS of *HFE* hemochromatosis

Marie de Tayrac[1,2,3], Marie-Paule Roth[4,5,6], Anne-Marie Jouanolle[3], Hélène Coppin[4,5,6], Gérald le Gac[7,8], Alberto Piperno[9], Claude Férec[7,8], Sara Pelucchi[9], Virginie Scotet[7,8], Edouard Bardou-Jacquet[10,11], Martine Ropert[12], Régis Bouvet[3,13], Emmanuelle Génin[7], Jean Mosser[1,2,3]*, Yves Deugnier[10,11,14]*

* Equally contributing contributors

1- CNRS, UMR 6290, Institut Génétique et Développement de Rennes, F-35043 Rennes, France

2- Université de Rennes 1, UEB, Biosit, Faculté de Médecine, F-35043 Rennes, France

3- CHU Rennes, Service de Génétique Moléculaire et Génomique, F-35033 Rennes, France

4- Inserm, U1043, F31024 Toulouse, France

5- CNRS, U5282, F31024 Toulouse, France

6- Université de Toulouse, UPS, Centre de Physiopathologie de Toulouse Purpan (CPTP), F31062 Toulouse, France

7- Inserm UMR1078, Université de Brest, SFR ScInBioS, CHRU, F-29200 Brest, France

8- Etablissement Français du Sang – Bretagne, Laboratoire de Génétique Moléculaire et d'Histocompatibilité, F-29200 Brest, France

9- University of Milano-Bicocca, Department of Health Sciences, Centre for Iron Disorders, S. Gerardo Hospital, Monza, Italy

10- CHU de Rennes, Liver Unit and French Reference Centre for Rare Iron Overload Diseases of Genetic Origin, F-35033 Rennes, France

11- INSERM, UMR 991, Equipe Fer et Foie, F-35033 Rennes, France

12- CHU Rennes, Service de Biochimie, F-35033 Rennes, France

13- Plate-forme Génomique Santé Biogenouest, Biosit, F-35043 Rennes, France

14- INSERM, Centre of Clinical Investigation 0203, CHU de Rennes, F-35033 Rennes cedex

Correspondence

Marie de Tayrac

CNRS UMR6290 Institut de Génétique et Développement de Rennes

Service de Génétique Moléculaire et Génomique

2, avenue Henri Le Guilloux

35033 Rennes cedex 9

France

Marie.detayrac@univ-rennes1.fr

Tel: +33 (0)2 99 28 99 70

Fax: +33 (0)2 99 28 92 48

Word count: 4123

Number of Figures: 2

Number of tables: 2

List of abbreviations

HH: Hereditary hemochromatosis

GWAS: genome-wide association study

SNP: Single Nucleotide Polymorphism

HAS: Haute Autorité de santé

Conflict of interest: None

Author's Contributions

Abstract

Background. Hereditary hemochromatosis (HH) is the most common form of genetic iron loading disease. It is mainly related to the *HFE* C282Y/C282Y homozygous genotype that is however a necessary but not a sufficient condition to develop clinical and even biochemical HH. This suggests that modifiers genes are likely involved in the expressivity of the disease. Our aim was to identify such modifier genes.

Patients and methods. We performed a genome-wide association study (GWAS) using DNAs collected from 474 unrelated C282Y homozygotes. Associations were examined for both quantitative iron burden indices and clinical outcomes with 534,213 single nucleotide polymorphisms (SNP) genotypes, with replication analyses in an independent sample of 748 C282Y homozygotes from four different European centres.

Results. One SNP met genome-wide statistical significance for association with transferrin concentration (rs3811647, GWAS p-value of 7×10^{-9} and replication p-value of 5×10^{-13}). This SNP located within intron 11 of *TF* gene had pleiotropic effect on serum iron (GWAS p-value of 4.9×10^{-6} and replication p-value of 3.2×10^{-6}). Both serum transferrin and iron levels were associated with serum ferritin levels, amount of iron removed and global clinical stage (p<0.01). Serum iron levels were also associated with fibrosis stage (p<0.0001).

Conclusion. This GWAS, the largest one performed so far in unselected *HFE*-HH patients, identified the rs3811647 polymorphism in the *TF* gene as the only SNP significantly associated with iron metabolism through serum transferrin and iron levels. Because these two outcomes were clearly associated with the biochemical and clinical expression of the disease, an indirect link between the rs3811647 polymorphism and the phenotypic presentation of HFE-HH is likely.

Keywords

HFE-hemochromatosis , GWAS, transferrin

Introduction

Genetic hemochromatoses are hereditary disorders proceeding from impairment in the production and/or activity of the key regulator of plasma iron, hepcidin, and resulting in progressive iron loading of parenchymas. *HFE*-hemochromatosis accounts for more than 95% of genetic hemochromatosis cases in Caucasian populations [1]. It is chiefly related to one major mutation, p.Cys282Tyr (C282Y), whose frequency ranges from 5% to 15% according to an increasing gradient from South/East to North/West but whose penetrance is far from complete [2].

Large studies based on systematic genotyping of large populations [2-6] have consistently found that transferrin saturation and serum ferritin levels, the key biochemical markers of genetic hemochromatosis, were significantly increased in C282Y homozygotes when compared to sex-matched controls with other *HFE* genotypes. According to studies, the percentage of C282Y homozygous subjects with elevated transferrin saturation ranged from 73 to 84% in males compared to 40 to 73% in females. The same was found for serum ferritin levels with a prevalence of increased values ranging from 70 to 88% in males and 33 to 69% in females. As a whole, the biochemical penetrance of C282Y in homozygotes can be estimated as about 75% in males and 50% in females. However, only a few studies of the clinical expressivity of C282Y homozygotes achieved a sufficient statistical power and referred to a control group to allow for a correct interpretation of the data. [2,5,7-9] Based on these studies, it can be concluded that the clinical expressivity of the C282Y homozygous genotype is indisputably much lower than previously admitted, averaging 28% in males and 1% in females when using strict definitions of both iron overload and iron-related organ damage [2]. The main manifestations of *HFE*-hemochromatosis are chronic fatigue, arthritis and liver disease. The frequency of heart disease and diabetes are not significantly increased in C282Y homozygotes compared with controls [1]. However fatigue and arthritis are not clearly correlated with the amount of iron burden. With respect to liver disease, serum ferritin levels lower than 1000µg/L allow exclusion of significant fibrosis, as demonstrated by Guyader et al [10] and confirmed by Allen et al [2]. Thus C282Y homozygosity appears to be a necessary but not a sufficient condition for the development of *HFE*-

hemochromatosis and the existence of both environmental and genetic modifiers of its penetrance is strongly suggested.

Although familial studies have shown that non-*HFE* iron-related genes are involved in the modulation of iron burden in C282Y homozygotes, only a few putative genetic modifiers have been identified so far in humans. A mitochondrial polymorphism has been reported as more frequent in iron loading expressing than in non-expressing C282Y homozygotes [11] but this was not confirmed by another study [12]. Co-inheritance of mutations in the hepcidin gene (*HAMP*) [13,14] or in the hemojuvelin gene (*HFE2*) [15] together with C282Y homozygosity was found to be associated with more severe iron burden of earlier onset in a French cohort of C282Y homozygotes [13,15]. However these mutations have a frequency lower than 2‰ among C282Y homozygotes and can thus explain increased disease severity in a few patients only. Based on these results, the BMP - SMAD pathway regulating hepcidin production was highlighted as an important pathway that was likely including potential modifying genes. This was confirmed by the finding, in a large cohort of C282Y homozygotes, of an association between common variants in the BMP pathway and the level of iron burden [16,17]. The full expression of *HFE*-hemochromatosis may thus be linked to abnormal hepatic expression of hepcidin, through both impairment in HFE function and functional modulation in the BMP pathway. Other results support the hypothesis that polymorphisms in different genes regulating iron metabolism may modulate penetrance of *HFE*-HH. *CYBRD1* was shown to be a possible modifier gene of iron phenotype in patients with *HFE*-HH [18,19] and, recently, the p.Ala736Val *TMPRSS6* polymorphism was shown to be likely a modifier of *HFE*-HH in a highly selected subgroup of male C282Y homozygotes [20].

We hypothesized that other common variants located in genes outside this pathway could modulate iron overload in *HFE*-HH. To identify these variants, we conducted a genome-wide association study in 518 C282Y homozygous *HFE*-HH patients from the tertiary referral centre of Rennes, France. The association between SNPs and 7 clinical and biochemical outcomes was examined and a replication study of candidate SNPs was conducted in a sample of 748 *HFE*-HH patients from 4 Italian and French centres. In this paper, we present the results of this GWAS study that is the largest conducted so far in *HFE*-HH patients.

Materials and Methods

Study population and replication sample

Approval for the genome-wide association study was obtained from the local ethical committee (reference 09/10-722 – 23/03/2009). All participants gave written informed consent. Samples were collected thanks to the Logifer database comprising all C282Y homozygous probands and relatives referred to the Liver Unit in Rennes, France, and collected retrospectively before 1989 and, then, prospectively. According to French rules, Logifer was declared to the national committee "Informatics and Freedom" (CNIL). Data are recorded from patient's files by data-managers and keyboarded twice. A clinical research assistant provides quality control, and coherence control and statistical analysis are performed by a bio-statistician.

Selection criteria included availability of sex, age at diagnosis, serum ferritin level at diagnosis before any venesection therapy, amount of iron removed by phlebotomy and frozen DNA samples stored at -20°C in the Centre of Biological Resources of Rennes. Carriers of the H63D mutation were excluded from this study. Of the 1730 C282Y homozygotes recorded in Logifer at the time of the study, 518 fulfilled these criteria. Replication for SNPs identified from the GWAS was conducted in 748 C282Y homozygous patients with informed consent from four different European Centres: Toulouse in Southern France (n=279), Brest and Rennes in Brittany (n=226 and n=124, respectively), and Monza in Northern Italy (n=119). Seven clinical and biochemical outcomes were examined: serum transferrin concentration, serum iron, serum ferritin, transferrin saturation, amount of iron removed by phlebotomy, liver fibrosis, and mixed biochemical and clinical score referred as the HAS score (Haute Autorité de Santé) [21]. All clinical and biochemical data were recorded at the time of diagnosis before any venesection therapy. Table 1 presents the characteristics of all the samples included in the GWAS and replication studies.

Liver fibrosis was assessed using the Metavir scoring system when liver biopsy was available (n=227) [22]. Patients were classified as having severe fibrosis when the Metavir score was either F3 or F4. In the absence of liver biopsy, according to Guyader's algorithm, severe liver fibrosis was considered as absent when serum

ferritin was < 1000 µg/L. HAS score [21] ranged from 0 to 4 according to the absence/presence of increased transferrin saturation (stage 1), elevated serum ferritin levels (stage 2), iron-related functional symptoms (asthenia, distal arthralgia; stage 3) and iron-related organ damage (liver fibrosis, diabetes; stage 4). In addition, the following data, recorded at the time of diagnosis, were available for the Rennes GWAS cohort: alcohol consumption, tobacco smoking, medicinal iron and drug intake, year(s) and number of blood donations, year(s) and number of pregnancies, age at menopause, weight and height, blood pressure, common blood tests including serum cholesterol, triglycerides, ASAT, ALAT, glucose, uric acid, creatinine, and HCV antibody and HBs antigen (negative in all patients).

Laboratory Methods

We determined iron status by measuring the levels of serum iron (mmol/L), transferrin saturation with iron (%) and serum ferritin (µg/L). Amount of iron removed (g) was assessed from the volume of blood extracted during initial depletion therapy until the achievement of low body iron stores (serum ferritin ≤ 50), assuming that 500 ml of blood contains 250 mg of elemental iron. For patients that were not subject to phlebotomy because of the absence of significant iron overload, it was assumed to be 0. Genomic DNA was extracted from peripheral blood cells by the phenol-chloroform method or by use of the Flexigen DNA kit (Qiagen). Only DNA samples with quality in accordance with Illumina recommendations were kept.

Genome-wide genotyping and quality control procedures

GWAS was performed on the 518 cases with the Illumina™ InfiniumHD technology (Human660W-Quad). A total of 657,366 markers were assessed for quality. Quality control tests were carried out using the GenABEL library [23] of the R statistical package (http://www.r-project.org/). The quality control flow diagram is presented in supplementary Fig. 1. Quality control assessments resulted in the exclusion of 95,988 CNV markers and 27,165 SNPs because of at least one of the following criteria: a minor allele frequency less than 1%, a call rate less than 95%, or rejection of Hardy-Weinberg equilibrium (HWE) at p-value equal to 10^{-6}. Thirty-one samples were excluded from the analysis due to a call rate <99%, too high autosomal

heterozygosity (FDR <1%), an average identity-by-state (IBS) value > 0.2, or misclassification of sex. Iterative principal component analyses (PCA) of a matrix of IBS distances computed from the SNP genotype data excluded 13 patients as heterogeneous in ancestry or outliers. After filtering of SNPs and samples, there were 534,213 SNPs available for analysis in a sample of 474 cases. Genetic population structure was assessed by PCA thanks to the OriginMineR R package. [24]

Genotyping and quality control procedures for replication samples

Replication for SNPs identified from the GWAS was carried out by custom Fluidigm SNPtype assays based on allele-specific PCR SNP detection chemistry and designed for the BioMark HD System. Each sample was assessed for completeness of data (>90%) and genotype data were assessed for deviation from HWE ($p<10^{-6}$), minor allele frequency (>5%) and completeness (>99%).

Statistical analyses

Statistical analyses of the GWAS genotype data were carried out using the GenABEL R library for genome-wide association analysis [23]. The quantitative iron status outcomes were analysed using linear regression after rank-based inverse normal transformation to avoid inflated type I error resulting from departure from normality. Genotypes were coded as 0, 1 or 2, indicating the number of copies of the less frequent of the two alleles in the genotype. The regression models for all outcomes included the additive genotype term and the covariates: age, with consideration of age groups of 10 years, and sex. Population stratification was also taken into account by adjusting models for the two principal components obtained from the PCA of the whole genotype data. For liver fibrosis, logistic regression was performed and we used ordered logistic regression (R cran function, polr) to model HAS-score. For the replication phase, a four-level factor indicating the centre where the sample was collected was added to each model instead of the principal components. As we conducted several analyses using different phenotypes and performed 8 * 534,213

tests in our entire GWA analysis, we considered p-values < 10^{-8} ($0.05/4.3 \times 10^{6}$) to indicate genome-wide significance in our study.

The 20 SNPs tested in the replication study consisted of (i) SNPs with $P<10^{-6}$ from which only two SNPs were selected by LD block (Table 2), and, (ii) SNPs with $P<10^{-5}$ and $P>10^{-6}$ with one SNP being selected for each LD block to avoid redundancy (Supplementary Table 1).

Genotype imputation

Data were imputed for over 10 million SNPs using data from the 1000 Genomes Project [25]. Genotypes were aligned to the positive strand in both imputation and genotyping. SHAPEIT2 [26] and IMPUTE2 [27] were used for imputation of non-genotyped SNPs in the Illumina Human660W-Quad chips. We removed SNPs with minor allele frequency less than 1% and impute info measure less than 0.8. Genome-wide association analyses were conducted using SNPtest [27].

Results

Characteristics of patients

The clinical and biological characteristics of patients belonging to the primary and replication cohorts are reported in Table 1. With respect to primary cohort, additional descriptive data were as follows: BMI (men: 25.15±3.27, women: 24.25± 4.5), excessive alcohol consumption (men: 16%, women: 5%), tobacco consumption (men: 31%, women:15%), blood donation (men: 23%, women: 21%), proportion of pregnancies (none=17%, one or 2: 37%, greater than 3: 24.5%) , number of menopausal women (44%). Link between biological iron data and these putative modifying factors was investigated using both univariate and multivariate Gaussian regression models. Alcohol consumption was strongly associated with body iron ($p=0.004$) and serum ferritin ($p=6\times10^{-7}$) in univariate analyses. This association remains significant only with serum ferritin in a multivariate analysis including gender and age. In women, pregnancy and menopause were associated with body iron ($p=0.04$ and $p=0.02$, respectively), serum ferritin ($p=0.04$ and $p=0.01$, respectively) and serum iron (NS and $p=0.002$, respectively) in univariate analyses only. Thus, GWAS was performed with regression including age and sex as only covariates. For ferritin, an analysis including also alcohol consumption was performed.

Genome-wide Association Study

The sequencing of the *HFE* gene confirmed that all participants were homozygous for the p.Cys282Tyr mutation and did not carry any other mutation in this gene known to be involved in HH. Following quality control, a total of 534,213 SNP and 474 patients were included in the GWAS (supplementary Fig. 1). All *HFE*-HH patients were identified as being from European ancestry and particularly from Western Europe (supplementary Fig. 2).

Figure 1 shows the Manhattan plot obtained for GWAS of all the seven clinical and biochemical outcomes. P-values $<10^{-6}$ were obtained for ten SNPs on chromosomes 2, 3, 6, 16 and 9 (Table 2) and 41 SNPs were found associated with at least one of the seven outcomes with p-values between 10^{-6} and 10^{-5} (Supplementary Table 1 and Supplementary Fig. 3). However, GWAS analyses identified only one locus reaching the genome-wide significance threshold of 10^{-8} (Figure 2). This locus associated with serum transferrin concentration was located on chromosome 3q22

(*TF* gene; rs3811647, intron 11, p-value = 7×10^{-9}). This SNP was also associated with serum iron (p-value = 4.9×10^{-6}) but not associated with other phenotypic descriptors. The effect of rs3811647 was additive with each A allele increasing both serum transferrin and serum iron concentration. It explained 7.7% of the variance of serum transferrin concentration and 4.7% of the variance of serum iron levels. Other potential associations were found between serum ferritin and rs3849200 on chromosome 6p16.3 (p-value = 6.2×10^{-7}) and rs4544229 on chromosome 16q21 (p-value = 8.9×10^{-7}), between the amount of iron removed and rs4371141 (*VAV2* gene) on chromosome 9q34 (p-value = 9.4×10^{-7}), and between the HAS-score and rs3900546 located between *LOC644013* and *MAP4K4* on chromosome 2 (p-value = 2×10^{-7}). Associations between serum ferritin and rs3849200 and rs4544229 were conserved when alcohol consumption was included in the regression model.

Follow-up Association Study

In the second phase, twenty SNPs were selected for validation on an independent sample of 748 patients. Iron parameters distributions for patients included in the replication study are presented Table 1. Data distributions were compared between centres with linear models adjusted by age and gender (Supplementary Fig. 4). Raw iron parameters were found to vary significantly between centres (Supplementary Fig. 4 and 5A) but differences between centres were clearly reduced after rank-based inverse normal transformations (Supplementary Fig. 5B). After quality controls, a total of 18 SNPs were retained for replication, rs3849200 and rs2810493 being removed because of no call rate. Among these SNPs, only rs3811647 in the *TF* gene showed evidence for association with both serum transferrin and serum iron with p-values of 5×10^{-13} and 3.6×10^{-6}, respectively (Supplementary Table 2).

Association Analysis of Combined GWAS and Follow-up Samples

The GWAS and follow-up study data were combined (n=1220 patients) and were analysed for each outcome in a model that included age, sex, the collection centre variable and a variable to indicate if patients were from the GWAS or the follow-up study. Results are presented in Supplementary Table 3. Genome-wide significance was found for the association of rs3811647 in the *TF* gene with serum transferrin (p-

value = 1×10^{-20}) and serum iron (p-value = 9.24×10^{-11}). Besides these strong associations, no significant direct association with other phenotypic descriptors was found. Four other loci show potential association signals but with low significance (p-value around 10^{-3}): rs10788066, located between the *PPAPDC1A* and *SEC23IP* genes, with serum ferritin, amount of iron removed, severe liver fibrosis, and HAS-score; rs2941832, near the gene *SALL3,* with serum ferritin and amount of iron removed; rs2810473 and rs4406471 in the VAV2 gene with serum ferritin and, to a lesser extent, amount of iron removed, and rs17012214, located within the gene *TTC27,* with severe liver fibrosis*.*

As the role of *TMPRSS6* has been highly documented in iron metabolism in the last few years [29-33], we also assessed, in the combined cohort, the association of the 7 outcomes with two common variants in *TMPRSS6* (rs855791 and rs4820268). Using the same models, we found weak signals of association with the HAS-score for both SNPs (rs855791, p-value = 1×10^{-4} and rs4820268, p-value = 6×10^{-4}), and with the amount of iron removed for rs4820268 only (p-value = 4×10^{-4}). No association was detected with other parameters of iron status.

rs3811647 genotypes in the general population
We investigate the frequency of rs3811647 alleles from genotyping data obtained in a non-HFE population from Brittany (n=191). We found that A allele has frequency of 0.343 (AA, n=21; AG, n=89; GG, n=81). We compared allele frequency in the population of C282Y homozygous patients against this population of Brittany and against GBR and CEU from 1000 Genomes and did not find any difference (chi-squared test).

Clinical, biochemical, and rs3811647 genotype correlations

Clinical and biochemical parameters were investigated in relation with rs3811647 genotypes in patients included in the GWAS study. Except for the additive effect of rs3811647, with each A allele increasing both serum transferrin and serum iron concentration, no significant influence of the rs3811647 genotype on clinical and other biochemical parameters was found (Table 3a). However, significant positive associations were detected between serum transferrin and serum iron levels and

between serum iron levels and all phenotypic outcomes studied. Serum transferrin levels were negatively correlated with all other phenotypic markers except fibrosis (Table 3b).

Discussion

In this paper, we report the first study designed at the genome-wide scale to identify SNPs associated with clinical and biochemical expression of *HFE*-hemochromatosis. This study was performed in a large sample of well-defined C282Y homozygous patients, totalizing more than 1,200 individuals presenting with a wide range of phenotypic expression. No significant association was found between SNPs and clinical outcomes. With respect to biochemical outcomes, only one chromosomal region including the rs3811647 SNP in the *TF* gene showed clear evidence for association with serum transferrin in the GWAS (p-value=7×10^{-9}), in the replication study (p-value=5×10^{-13}) and the combined analysis (p-value=1×10^{-20}). Unexpectedly, this SNP was also clearly associated with serum iron, and therefore indirectly linked to biochemical and clinical outcomes.

Notably, in the general population, rs3811647 was reported as associated with serum transferrin but not with serum iron [33-35] contrary to what we found in hemochromatosis patients. It is possible that, in the general population, the power of detecting association with serum iron was compromised by the limited range of the values of this parameter. In contrast, our sample of C282Y homozygotes is a combination of biochemically expressing and non-expressing individuals. The necessarily wider range of serum iron in these patients may have allowed us to highlight the role of rs3811647 genotype in modulating not only serum transferrin but also serum iron levels.

However, our study did not find a direct link between the rs3811647 SNP in the *TF* gene and the phenotypic presentation of HH. Indeed, although this SNP was associated with serum iron, it was not significantly associated with serum ferritin, transferrin saturation, or any of the clinical outcomes tested. Indirect tests of associations are less powered than direct tests and may require much larger sample sizes than the ones used in this study to allow us to demonstrate a possible influence of the genotype at rs3811647 on other parameters of disease severity. However, we cannot exclude that such an association exists in HFE hemochromatosis only

through the dysregulation of iron cellular efflux secondary to the impairment of hepcidin expression (Figure 3).

Our study indicates that the association of rs3811647 with serum transferrin concentration is maintained in C282Y homozygous patients and that the lowest serum transferrin levels are observed in those bearing the GG genotype. However, even the patients with the rs3811647 AA genotype have less serum transferrin than normal controls polymorphism (data not shown). This finding fits with the common observation of decreased serum transferrin levels in C282Y homozygotes, which is attributed to the preservation of a physiologic negative feedback of transferrin synthesis by elevated body iron stores. This could contribute to aggravate the iron burden through the increase of transferrin saturation and the consecutive increase of non-transferrin bound iron, which is considered as the main factor of visceral iron loading in hemochromatosis. However, although they had lower serum transferrin levels, patients with the rs3811647 GG genotype had no increased transferrin saturation when compared to other patients. This is in line with the concomitant decrease in serum iron levels, and could suggests that transferrin levels, as determined by rs3811647 polymorphisms, may participate *per se* in the regulation of systemic iron, possibly through a direct interaction with the molecular cascade involved in hepcidin production. [37]

We took advantage of this study to investigate putative candidate genes selected from literature and/or coding for molecules involved in iron-related metabolic pathways and previously identified GWAS-significant SNPs (Supplementary Table 4 and Supplementary Table 5). No association, even borderline, was found. Finally, we focused on *TMPRSS6*, a modifier gene identified by Valenti *et al.* in a highly selected population of 96 non-alcoholic male C282Y homozygotes [20]. In the subset of 230 non-alcoholic male C282Y homozygotes from the GWAS sample, an association was detected with SNPs located in this gene at the 1% nominal level (p-value = 0.006 with AIR), but this association did not resist multiple testing corrections at the genome-wide level. It remains possible that *TMPRSS6* acts as a modifier-gene but with a weak effect compared to age and sex. HH is a long-standing disease whose diagnosis is made at adulthood. Along the constitution of iron excess, many factors

are able to act, at various periods and in opposite directions, resulting in highly variable iron burden from an individual to another. These factors include, apart blood donation, pregnancies and menopause that are easy to take into account and to quantify, regimen, drugs (including iron auto-medication), menstrual blood loss, alcohol consumption and evolution of metabolic features that are difficult to reliably assess over a long period. This highlights the need for better controlling environmental factors, not only alcohol consumption, which implies to study much larger population than in the present GWAS.

The present study suffered from several limitations. First, the phenotypic description of HH remains difficult and imprecise. Two types of indicators are available: those measuring the iron burden and those assessing organ damage. The degree of iron burden is usually deduced, at the time of diagnosis, from serum ferritin levels, hepatic iron concentration assessed either by biochemical determination or magnetic resonance imaging, histological semi-quantitative evaluation of hepatic iron stores, and, to a lesser extent, transferrin saturation. However, all these indicators have to be interpreted according to age and gender and are subject to frequent limitations. Conditions that often coexist in hemochromatosis and may result in overestimation of iron burden can modify serum ferritin levels (i.e. excessive alcohol intake, metabolic syndrome, cell necrosis and inflammatory syndrome). Because it requires an invasive approach, histological and biochemical determination of hepatic iron concentration is less and less performed. It also depends on both liver volume and hepatic iron distribution that may be heterogeneous, especially in case of cirrhosis. Hepatic iron concentration measurement by MRI is reliable within the range of 60-300 μmol/g of dry liver only, which precludes correct assessment of iron burden in patients with early and late hemochromatosis. Finally, transferrin saturation is a rough indicator of iron burden because it is usually rapidly complete regardless of excess iron amount. Then, we chose to select only patients with amount of iron removed (AIR) available, since AIR is considered as the most reliable means to assess iron burden. However, because AIR is retrospective and often difficult to calculate, numerous C282Y homozygotes of the cohort were lost for inclusion. Indicators of organ damage allow for a precise case description, but the non-specificity of most clinical symptoms and high prevalence of co-damaging factors, such as metabolic syndrome and alcohol,

may result in overestimating the role of iron in the clinical presentation. Second, the present study may have been underpowered due to the limited number of cases for a GWAS (n = 474 from the 518 eligible cases) and to the few participants with low or normal body iron stores. Because the frequency of HFE C282Y homozygotes is very low, collecting enough sample size for such a study design to detect variants with moderate effect size is a challenge. Alternative strategy might be needed, such as conducting GWAS of iron status in the general population and then test the identified loci in the C282Y homozygous patients. Large genome-wide studies in which the authors looked for loci controlling serum iron, serum transferrin, transferrin saturation and serum ferritin in the general population have been published recently [33-35]. Although performed on large sample sizes, these studies did not provide evidence for variants with significant effects on any of the parameters examined, except for variants in the *HFE* and *TMPRSS6* genes that were controlling serum iron, and variants in *TF* that were controlling serum transferrin. In this study we found that, in addition to being associated with serum transferrin, the same variants in the *TF* gene were also associated with serum iron in hemochromatosis patients. If we consider less stringent p-values, the variants in *TMPRSS6* can also be considered as marginally associated with some of the iron burden parameters in this sample of hemochromatosis patients. The success rate of this strategy is therefore quite good, but there were not many variants segregating with iron burden in the general population. In addition, it is also possible that different variants control iron parameters in the general population and in hemochromatosis patients, which motivated this particular study. Unfortunately, the C282Y mutation is not routinely screened in the general population, and only few participants with low or normal parameters of iron burden are included in our GWAS study. Therefore, the power of this study is probably compromised to some extent by the limited range of expression of these parameters. Although large-scale studies in the general population could help identify such C282Y homozygotes with low or normal iron parameters, it should be remembered that the frequency of this genotype is only 3-5/1000, and obtaining a number of non-penetrant cases from the same area sufficient to increase the power of this study is unrealistic for the moment. The third limitation is the lack of suitable information about environmental factors. For example, data on regimen with respect to daily iron intake and tea consumption were missing in the database. However, data on alcohol consumption, tobacco smoking, blood donation, and, in females,

number of pregnancies and menopause status were collected and analyzed. We did not find significant association when taking alcohol consumption into account but samples sizes were probability too small to detect possible gene-environment interactions. Fourth, although a large number of SNPs were tested (n = 534,213), the present study was far from covering all putative genetic variants able to act as modifiers of iron burden and/or organ damage. Using the SNP-chip data, we investigated the possible role of CNV and did not find any association. However, we were not able to investigate low frequency and rare variants that are not targeted nor tagged by the SNP-chip. We thus cannot exclude based on our results the involvement of one or more modifier genes but we can rule out the role of common variants with large effect outside the *TF* gene.

In conclusion, our results indicate that, in C282Y/C282Y *HFE*-HH, the rs3811647 polymorphism in the *TF* gene continues to regulate serum transferrin levels and is unexpectedly associated with serum iron levels. Our study also demonstrates a potential indirect but weak link, through serum iron levels, between the rs3811647 polymorphism and the clinical and biological expression of the disease. Finally it shows, by contrast to candidate gene studies, that, in an unselected and large population of C282Y homozygotes, i.e. in "true life", there is no other common polymorphism exerting a significant phenotypic effect. At this stage, the role of rare and low frequency variants that are not tagged by SNP-chips remains to be investigated. Next-generation sequencing technologies will enable to explore their role and determine if there exist some rare mutations with clinical significance in HFE-HH. In designing exome or whole genome sequencing studies, it will however be important to carefully select patients and to control for environmental factors.

References

[1] Liver, E. A. F. T. S. O. T. (2010), 'EASL clinical practice guidelines for HFE hemochromatosis.', *J Hepatol* **53**(1), 3--22.

[2] Allen, K. J.; Gurrin, L. C.; Constantine, C. C.; Osborne, N. J.; Delatycki, M. B.; Nicoll, A. J.; McLaren, C. E.; Bahlo, M.; Nisselle, A. E.; Vulpe, C. D.; Anderson, G. J.; Southey, M. C.; Giles, G. G.; English, D. R.; Hopper, J. L.; Olynyk, J. K.; Powell, L. W. & Gertig, D. M. (2008), 'Iron-overload-related disease in HFE hereditary hemochromatosis.', *N Engl J Med* **358**(3), 221--230.

[3] Adams, L. A.; Angulo, P.; Abraham, S. C.; Torgerson, H. & Brandhagen, D. (2006), 'The effect of the metabolic syndrome, hepatic steatosis and steatohepatitis on liver fibrosis in hereditary hemochromatosis.', *Liver Int* **26**(3), 298--304.

[4] Beutler, E.; Felitti, V. J.; Koziol, J. A.; Ho, N. J. & Gelbart, T. (2002), 'Penetrance of 845G--> A (C282Y) HFE hereditary haemochromatosis mutation in the USA.', *Lancet* **359**(9302), 211--218.

[5] Deugnier, Y.; Jouanolle, A.; Chaperon, J.; Moirand, R.; Pithois, C.; Meyer, J.; Pouchard, M.; Lafraise, B.; Brigand, A.; Caserio-Schoenemann, C.; Mosser, J.; Adams, P.; Gall, J. L. & David, V. (2002), 'Gender-specific phenotypic expression and screening strategies in C282Y-linked haemochromatosis: a study of 9396 French people.', *Br J Haematol* **118**(4), 1170--1178.

[6] Jackson, H. A.; Carter, K.; Darke, C.; Guttridge, M. G.; Ravine, D.; Hutton, R. D.; Napier, J. A. & Worwood, M. (2001), 'HFE mutations, iron deficiency and overload in 10,500 blood donors.', *Br J Haematol* **114**(2), 474--484.

[7] Adams, P. C.; Reboussin, D. M.; Barton, J. C.; McLaren, C. E.; Eckfeldt, J. H.; McLaren, G. D.; Dawkins, F. W.; Acton, R. T.; Harris, E. L.; Gordeuk, V. R.; Leiendecker-Foster, C.; Speechley, M.; Snively, B. M.; Holup, J. L.; Thomson, E.; Sholinsky, P.; Hemochromatosis & Investigators, I. O. S. (H. S. R. (2005), 'Hemochromatosis and iron-overload screening in a racially diverse population.', *N Engl J Med* **352**(17), 1769--1778.

[8] Andersen, R. V.; Tybjaerg-Hansen, A.; Appleyard, M.; Birgens, H. & Nordestgaard, B. G. (2004), 'Hemochromatosis mutations in the general population: iron overload progression rate.', *Blood* **103**(8), 2914--2919.

[9] Beutler, E. (1997), 'The significance of the 187G (H63D) mutation in hemochromatosis.', *Am J Hum Genet* **61**(3), 762--764.

[10] Guyader, D.; Jacquelinet, C.; Moirand, R.; Turlin, B.; Mendler, M. H.; Chaperon, J.; David, V.; Brissot, P.; Adams, P. & Deugnier, Y. (1998), 'Noninvasive prediction of fibrosis in C282Y homozygous hemochromatosis.', *Gastroenterology* **115**(4), 929--936.

[11] Livesey, K. J.; Wimhurst, V. L. C.; Carter, K.; Worwood, M.; Cadet, E.; Rochette, J.; Roberts, A. G.; Pointon, J. J.; Merryweather-Clarke, A. T.; Bassett, M. L.; Jouanolle, A.; Mosser, A.; David, V.; Poulton, J. & Robson, K. J. H. (2004), 'The 16189 variant of mitochondrial DNA occurs more frequently in C282Y homozygotes with haemochromatosis than those without iron loading.', *J Med Genet* **41**(1), 6--10.

[12] Beutler, E.; Beutler, L.; Lee, P. L. & Barton, J. C. (2004), 'The mitochondrial nt 16189 polymorphism and hereditary hemochromatosis.', *Blood Cells Mol Dis* **33**(3), 344--345.

[13] Jacolot, S.; Gac, G. L.; Scotet, V.; Quere, I.; Mura, C. & Ferec, C. (2004), 'HAMP as a modifier gene that increases the phenotypic expression of the HFE pC282Y homozygous genotype.', *Blood* **103**(7), 2835--2840.

[14] Merryweather-Clarke, A. T.; Cadet, E.; Bomford, A.; Capron, D.; Viprakasit, V.; Miller, A.; McHugh, P. J.; Chapman, R. W.; Pointon, J. J.; Wimhurst, V. L. C.; Livesey, K. J.; Tanphaichitr, V.; Rochette, J. & Robson, K. J. H. (2003), 'Digenic inheritance of mutations in HAMP and HFE results in different types of haemochromatosis.', *Hum Mol Genet* **12**(17), 2241--2247.

[15] Le Gac, G.; Scotet, V.; Ka, C.; Gourlaouen, I.; Bryckaert, L.; Jacolot, S.; Mura, C. & Férec, C. (2004), 'The recently identified type 2A juvenile haemochromatosis gene (HJV), a second candidate modifier of the C282Y homozygous phenotype.', *Hum Mol Genet* **13**(17), 1913--1918.

[16] Milet, J.; Dehais, V.; Bourgain, C.; Jouanolle, A. M.; Mosser, A.; Perrin, M.; Morcet, J.; Brissot, P.; David, V.; Deugnier, Y. & Mosser, J. (2007), 'Common variants in the BMP2, BMP4, and HJV genes of the hepcidin regulation pathway modulate HFE hemochromatosis penetrance.', *Am J Hum Genet* **81**(4), 799--807.

[17] Milet, J.; Gac, G. L.; Scotet, V.; Gourlaouen, I.; Thèze, C.; Mosser, J.; Bourgain, C.; Deugnier, Y. & Férec, C. (2010), 'A common SNP near BMP2 is associated with severity of the iron burden in HFE p.C282Y homozygous patients: a follow-up study.', *Blood Cells Mol Dis* 44(1), 34--37.

[18] Constantine, C. C.; Anderson, G. J.; Vulpe, C. D.; McLaren, C. E.; Bahlo, M.; Yeap, H. L.; Gertig, D. M.; Osborne, N. J.; Bertalli, N. A.; Beckman, K. B.; Chen, V.; Matak, P.; McKie, A. T.; Delatycki, M. B.; Olynyk, J. K.; English, D. R.; Southey, M. C.; Giles, G. G.; Hopper, J. L.; Allen, K. J. & Gurrin, L. C. (2009), 'A novel association between a SNP in CYBRD1 and serum ferritin levels in a cohort study of HFE hereditary haemochromatosis.', *Br J Haematol* 147(1), 140--149.

[19] Pelucchi, S.; Mariani, R.; Calza, S.; Fracanzani, A. L.; Modignani, G. L.; Bertola, F.; Busti, F.; Trombini, P.; Fraquelli, M.; Forni, G. L.; Girelli, D.; Fargion, S.; Specchia, C. & Piperno, A. (2012), 'CYBRD1 as a modifier gene that modulates iron phenotype in HFE p.C282Y homozygous patients.', *Haematologica* 97(12), 1818--1825.

[20] Valenti, L.; Fracanzani, A. L.; Rametta, R.; Fraquelli, M.; Soverini, G.; Pelusi, S.; Dongiovanni, P.; Conte, D. & Fargion, S. (2012), 'Effect of the A736V TMPRSS6 polymorphism on the penetrance and clinical expression of hereditary hemochromatosis.', *J Hepatol* **57**(6), 1319--1325.

[21] Brissot, P.; Troadec, M.; Bardou-Jacquet, E.; Lan, C. L.; Jouanolle, A.; Deugnier, Y. & Loréal, O. (2008), 'Current approach to hemochromatosis.', *Blood Rev* **22**(4), 195--210.

[22] Bedossa, P. & Poynard, T. (1996), 'An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group.', *Hepatology* **24**(2), 289--293.

[23] Aulchenko, Y. S.; Ripke, S.; Isaacs, A. & van Duijn, C. M. (2007), 'GenABEL: an R library for genome-wide association analysis.', *Bioinformatics* **23**(10), 1294--1296.

[24] de Tayrac, M.; Babron M.C.; Génin E. (2012) 'OriginMiner : show me your genes and I'll tell you where you come from.' Presented at the 62nd Annual Meeting of The American Society of Human Genetics, November 7, 2012 in San Francisco, California.

[25] The 1000 Genomes Project Consortium. (2010). 'A map of human genome variation from population-scale sequencing.' *Nature* **467** (7319):1061-1073.

[26] O. Delaneau, J. Marchini, 'The 1000 Genomes Project Consortium (2014) Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel'. *Nature Communications* **5** 3934

[27] Marchini J, Howie B, Myers S, McVean G, Donnelly P. (2007) 'A new multipoint method for genome-wide association studies by imputation of genotypes.' *Nat Genet* **39**(7):906–13.

[28] Pruim, R. J.; Welch, R. P.; Sanna, S.; Teslovich, T. M.; Chines, P. S.; Gliedt, T. P.; Boehnke, M.; Abecasis, G. R. & Willer, C. J. (2010), 'LocusZoom: regional visualization of genome-wide association scan results.', *Bioinformatics* **26**(18), 2336--2337.

[29] Finberg, K. E.; Heeney, M. M.; Campagna, D. R.; Aydinok, Y.; Pearson, H. A.; Hartman, K. R.; Mayo, M. M.; Samuel, S. M.; Strouse, J. J.; Markianos, K.; Andrews, N. C. & Fleming, M. D. (2008), 'Mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA).', *Nat Genet* **40**(5), 569--571.

[30] Du, X.; She, E.; Gelbart, T.; Truksa, J.; Lee, P.; Xia, Y.; Khovananth, K.; Mudd, S.; Mann, N.; Moresco, E. M. Y.; Beutler, E. & Beutler, B. (2008), 'The serine protease TMPRSS6 is required to sense iron deficiency.', *Science* **320**(5879), 1088--1092.

[31] Chambers, J. C.; Zhang, W.; Li, Y.; Sehmi, J.; Wass, M. N.; Zabaneh, D.; Hoggart, C.; Bayele, H.; McCarthy, M. I.; Peltonen, L.; Freimer, N. B.; Srai, S. K.; Maxwell, P. H.; Sternberg, M. J. E.; Ruokonen, A.; Abecasis, G.; Jarvelin, M.; Scott, J.; Elliott, P. & Kooner, J. S. (2009), 'Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels.', *Nat Genet* **41**(11), 1170--1172.

[32] Benyamin, B.; Ferreira, M. A. R.; Willemsen, G.; Gordon, S.; Middelberg, R. P. S.; McEvoy, B. P.; Hottenga, J.; Henders, A. K.; Campbell, M. J.; Wallace, L.; Frazer, I. H.; Heath, A. C.; de Geus, E. J. C.; Nyholt, D. R.; Visscher, P. M.; Penninx, B. W.; Boomsma, D. I.; Martin, N. G.; Montgomery, G. W. & Whitfield, J. B. (2009), 'Common variants in TMPRSS6 are associated with iron status and erythrocyte volume.', *Nat Genet* **41**(11), 1173--1175.

[33] Tanaka, T.; Roy, C. N.; Yao, W.; Matteini, A.; Semba, R. D.; Arking, D.; Walston, J. D.; Fried, L. P.; Singleton, A.; Guralnik, J.; Abecasis, G. R.; Bandinelli, S.; Longo, D. L. & Ferrucci, L. (2010), 'A genome-wide association analysis of serum iron concentrations.', *Blood* **115**(1), 94--96.

[34] Benyamin, B.; McRae, A. F.; Zhu, G.; Gordon, S.; Henders, A. K.; Palotie, A.; Peltonen, L.; Martin, N. G.; Montgomery, G. W.; Whitfield, J. B. & Visscher, P. M. (2009), 'Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels.', *Am J Hum Genet* **84**(1), 60--65.

[35] Pichler, I.; Minelli, C.; Sanna, S.; Tanaka, T.; Schwienbacher, C.; Naitza, S.; Porcu, E.; Pattaro, C.; Busonero, F.; Zanon, A.; Maschio, A.; Melville, S. A.; Piras, M. G.; Longo, D. L.; Guralnik, J.; Hernandez, D.; Bandinelli, S.; Aigner, E.; Murphy, A. T.; Wroblewski, V.; Marroni, F.; Theurl, I.; Gnewuch, C.; Schadt, E.; Mitterer, M.; Schlessinger, D.; Ferrucci, L.; Witcher, D. R.; Hicks, A. A.; Weiss, G.; Uda, M. & Pramstaller, P. P. (2011), 'Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels.', *Hum Mol Genet* **20**(6), 1232--1240.

[36] Ruchala P, Nemeth E. (2014) 'The pathophysiology and pharmacology of hepcidin.' *Trends Pharmacol Sci.* **35**(3):155-61.

[37] Finberg, K. E. (2013), 'Regulation of systemic iron homeostasis.', Curr Opin Hematol 20(3), 208--214.

Tables

Table 1 – Main clinical and demographic characteristics of the primary and replication cohorts

	GWAS sample	Replication samples			
		Rennes	Toulouse	Brest	Monza
Sample size	474	124	279	224	119
Gender (female) (male)	(200)(274)	(67)(57)	(108)(171)	(93)(133)	(26)(93)
Age at diagnosis (years)	45±12	44±14.2	47±12	46.0±13	45±12.2
Serum iron (mmol/l)	34.5 (29 - 39.3)	34.00 (29.34 - 38.90)	35.6 (30.4 - 40.8)	36.0 (31.0 - 41.0)	36.43(33.48 - 39.55)
NA's	26	11	44	16	0
Serum Transferrin (g/L)	1.80 (1.60-2.00)	1.88 (1.60-2.01)	1.78 (1.60-2.03)	1.77 (1.60-2.00)	1.83 (1.58-2.06)
NA's	38	14	55	16	0
Transferrin saturation (%)	80 (65 - 89)	78 (66.5 - 85.85)	84 (67 - 92)	83 (73 - 92)	84 (72 - 90)
NA's	28	10	35	0	33
AIR (g)	4.7 (2.1 - 8.2)	3.8 (1.7 - 7.8)	6.1 (3.5 - 11.2)	4.1 (2.12 - 6.8)	6(4 - 10.75)
NA's	0	0	197	0	0
Serum ferritin (µg/L)	776 (446 - 1565)	570 (318 - 1136)	1118 (634 - 2315)	688 (370 - 995)	1120 (719-2374)
NA's	11	5	0	0	0
Fibrosis (grading)					
no	367	100	128	161	81
severe	72	14	30	41	25
NA's	35	10	121	22	13
HAS (grading)					
0	13	3	3	0	1
1	44	12	10	0	3
2	131	45	89	0	48
3	164	32	141	0	39
4	76	18	36	0	28
NA's	46	14	0	224	0

a Median value ± sd.

b Median value (25th–75th percentile).
* Amount of Iron Removed (weekly or bimonthly venesections until serum ferritin < 50; 1 L of blood = 0,5 g of iron)
NA = non available

Table 2. GWA Analyses results with P-values <10^{-6}

SNP	Trait	Chr	Position	Closest gene	Location	A1	A2	maf	N	effB	Pvalue	Geno1.freq	Geno2.freq	Geno3.freq	Replication
rs2692695	sTF	3	134968144	TF	intron	G	A	0.41	437	-0.35	1.85E-07	A/A: 78	G/A:230	G/G:166	NA
rs3811647	**sTF**	**3**	**134966719**	**TF**	**intron**	**G**	**A**	**0.33**	**437**	**0.41**	**7.00E-09**	**A/A: 46**	**G/A:217**	**G/G:211**	**X**
rs6794945	sTF	3	135001153	SRPRB	flanking_5UTR	G	A	0.3	437	0.38	6.03E-07	A/A: 33	G/A:219	G/G:222	NA
rs3849200	sFRT	6	101765612	GRIK2	flanking_5UTR	G	A	0.1	463	0.44	6.19E-07	A/A: 5	G/A: 88	G/G:381	X
rs4371141	sFRT	16	57773110	GOT2	flanking_5UTR	G	A	0.19	463	0.34	9.43E-07	A/A: 18	G/A:144	G/G:312	X
rs4544229	sFRT	16	57786093	GOT2	flanking_5UTR	G	A	0.19	463	0.34	8.91E-07	A/A: 18	G/A:143	G/G:313	X
rs2810472	AIR	9	135750460	VAV2	intron	A	G	0.25	472	-0.31	9.79E-07	A/A:271	A/G:170	G/G: 33	NA
rs2810473	AIR	9	135750672	VAV2	intron	G	A	0.25	472	-0.31	9.79E-07	A/A: 33	G/A:170	G/G:271	X
rs2810493	AIR	9	135756956	VAV2	intron	G	A	0.27	472	-0.3	9.44E-07	A/A: 36	G/A:181	G/G:257	X
rs3900546	HAS	2	101560458	FLJ42986	flanking_5UTR	A	C	0.26	474	-0.75	2.00E-07	A/A:265	A/C:170	C/C: 39	X

Abbreviations: sTF=serum transferrin concentration, sFRT= serum ferritin concentration, AIR= amount of iron removed, HAS= HAS score, A1= reference allele, A2: non-reference allele, maf=minor allele frequency, effB= Effect of the non-reference allele in allelic test.

Table 3. rs3811647 genotypes, clinical and biochemical correlations

a.

rs3811647	Serum iron	Serum transferrin	Transferrin saturation	Serum ferritin	Iron removed	Fibrosis	HAS score
AA	**38.8 ± 9.1** (n=43)	**1.98 ± 0.30** (n=43)	78.3 ± 17.4 (n=44)	1632 ± 1646 (n=45)	7.3 ± 6.2 (n=45)	0.18 ± 0.38 (n=40)	2.66 ± 1.03 (n=44)
AG	**34.3 ± 9.1** (n=207)	**1.87 ± 0.28** (n=198)	74.7 ± 18.8 (n=203)	1145 ± 1329 (n=213)	5.9 ± 4.9 (n=210)	0.14 ± 0.35 (n=198)	2.56 ± 1.03 (n=201)
GG	**32.4 ± 7.1** (n=98)	**1.74 ± 0.33** (n=195)	75.9 ± 17.8 (n=199)	1296 ± 1356 (n=205)	6.7 ± 6.1 (n=201)	0.16 ± 0.36 (n=194)	2.56 ± 0.99 (n=200)
	p<0.001	**p<0.001**	p=0.453	p=0.08	p=0.188	p=0.844	p=0.837

b.

Phenotypic markers	Serum iron	Serum transferrin	Transferrin saturation	Serum ferritin	Iron removed	Fibrosis	HAS score
Serum iron		**r = 0.257 (n=436) p<0.0001**	**r = 0.747 (n=437) p<0.0001**	**r = 0.279 (n=445) p<0.0001**	**r = 0.216 (n=430) p<0.0001**	**r = 0.165 (n=411) p=0.0008**	**r = 0.346 (n=424) p<0.0001**
Serum transferrin	**r = 0.257 (n=436) p<0.0001**		**r = - 0.388 (n=436) p<0.0001**	**r = - 0.081 (n=433) p<0.0001**	**r = -0.123 (n=419) p=0.0118**	r = - 0.025 (n=402) p=0.6115	**r = - 0.126 (n=414) p=0.0104**
Transferrin saturation	**r = 0.747 (n=437) p<0.0001**	**r = - 0.388 (n=436) p<0.0001**		**r = 0.347 (n=443) p<0.0001**	**r = 0.339 (n=429) p<0.0001**	**r = 0.197 (n=410) p=0.0001**	**r = 0.437 (n=424) p<0.0001**

Correlations between semi-quantitative variables (fibrosis and HAS score) and quantitative variables were also performed using one-way analysis of variance and gave the same results.

Figure Legends

Figure 1 - Manhattan plots displaying results from GWAS of quantitative iron burden indices and clinical outcomes of HFE-HH.

Genome-wide analyses were performed for quantitative iron burden indices (serum transferrin concentration, serum iron, transferrin saturation, amount of iron removed by phlebotomy, and serum ferritin), and clinical outcomes (severe liver fibrosis, and HAS-score). All the analyses were performed on 534,213 genotyped SNPs. Horizontal lines indicate a suggestive p-value threshold of 10^{-6}.

Figure 2 – Regional plots for rs3811647 located within *TF* gene.

Regional plots for imputed results are presented for rs3811647 identified as the single locus reaching the genome-wide significance threshold of 10^{-8}: for A. serum transferrin concentration analysis (p-value = 7×10^{-9}), and B. with pleiotropic effect on serum iron (p-value = 4.9×10^{-6}). The rs3811647 SNPs is represented as a diamond. Other SNPs are gray scaled following their r^2 value with the most associated SNP. The blue line represents variation in recombination rate. Plots were generated by LocusZoom. [28]

Figure 3 – Regulation of hepcidin transcription through the bone morphogenetic protein (BMP) – Smad pathway.

Binding of holotransferrin to transferrin receptor 1 (TfR1) displaces HFE from TfR1 to TfR2. The HFE-TfR2 complex binds to a complex associating BMPs, BMP receptor (BMPR) and hemojuvelin (HJV). Neogenin and transmembrane serine protease 6 (TMPRSS6) modulate hepcidin transcription by cleaving HJV. Hepcidin acts by favoring the internalization and degradation of the transmembrane iron transporter, ferroportin (FRP), which results in decreasing the systemic biodisponiblity of iron. In HFE hemochromatosis, the C282Y mutation impairs the migration of HFE to the cell membrane and thus the activation of the BMP – BMPR- HJV complex. This results in the preservation of FRP at the cell membrane and thus a high iron efflux. [36]
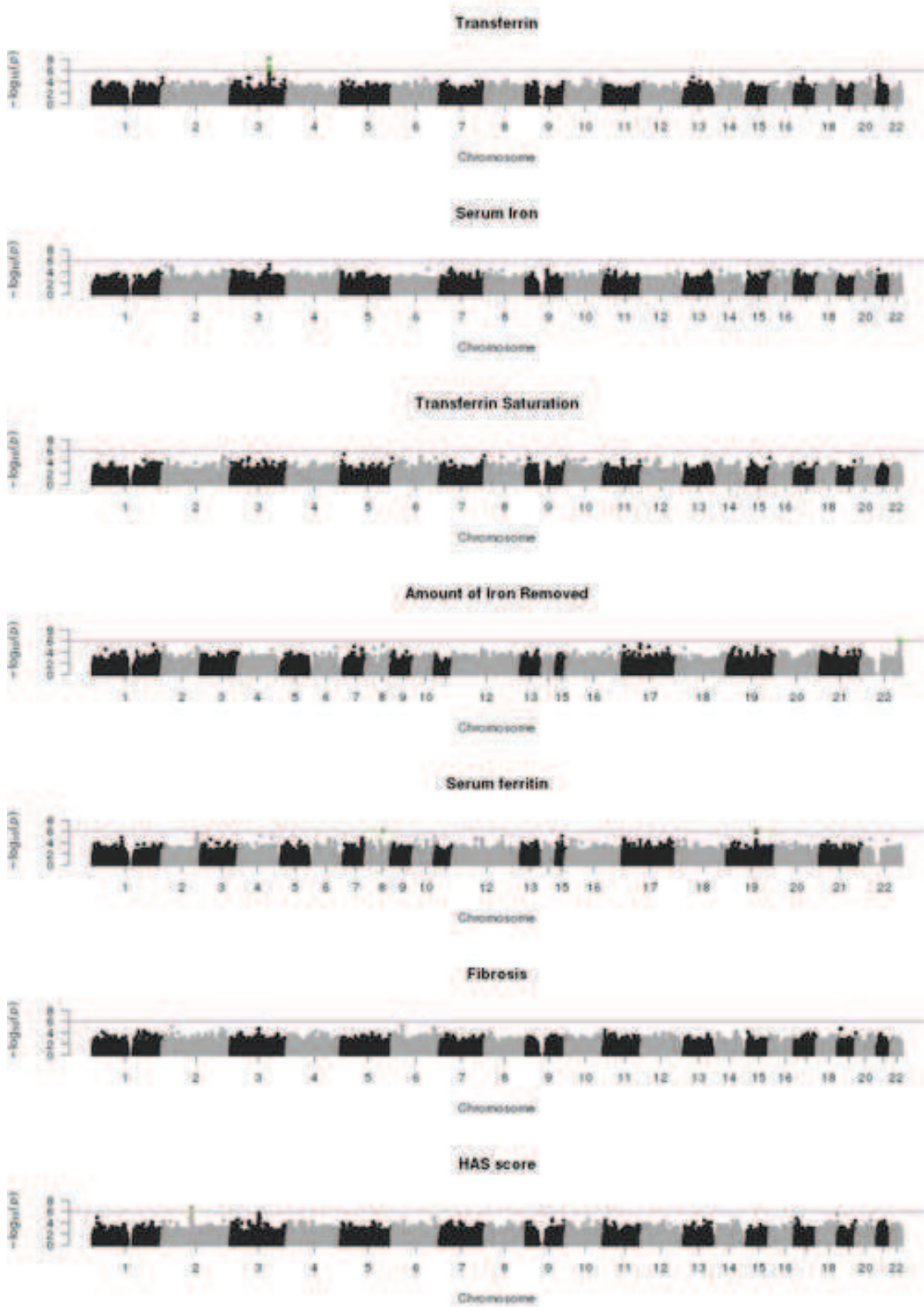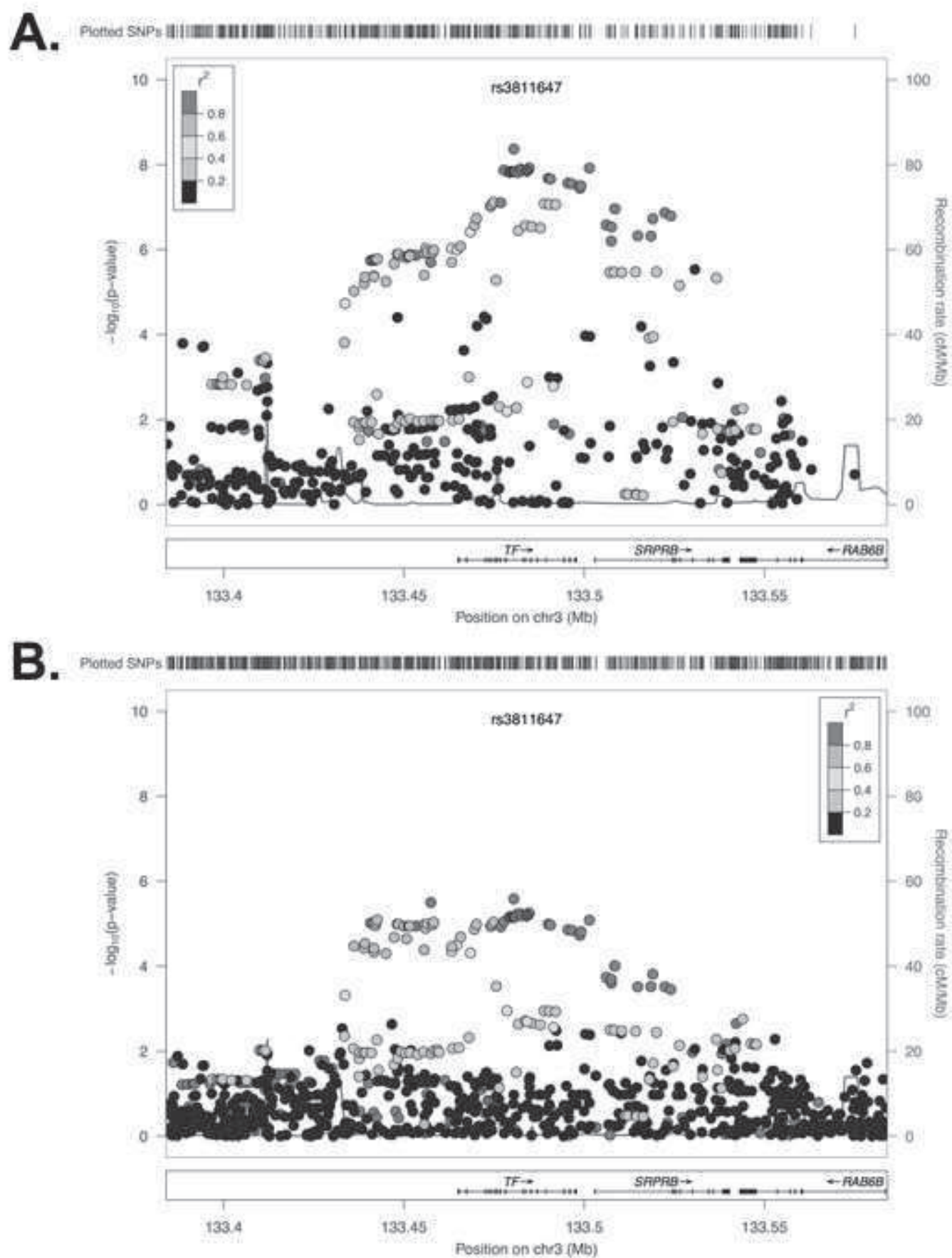
Figure 1

Figure 2

Figure 3